MÉTODOS PARA EL CÁLCULO DE LOS AUTOVALORES DE UNA MATRIZ

JUAN MANUEL PALOMAR ANGUAS

DNI: 07536788-X

Domicilio: C/ Juan de Juanes, nº 1, Esc. Izq. 8º C

C. P. 28933 Móstoles, MADRID

Teléfono: **91 614 25 57**

E-mail: **jmplmr@hotmail.com**

Departamento de Matemática Aplicada

Universidad de Alcalá de Henares

Resumen

En la presente exposición se van a describir todas aquellas ideas que aun siguen en vigor para resolver el cálculo de los autovalores de una matriz, contribuyendo en el desarrollo de potentes algoritmos numéricos.

Mediante la teoría de autovalores se posibilita la diagonalización de una matriz o la reducción a la forma de Jordan cuando la diagonalización no es posible. Dicho procedimiento facilita el análisis de un endomorfismo eligiendo de entre las matrices semejantes que lo representan, la más sencilla.

Algunas aplicaciones del cálculo de los autovalores se refieren a la resolución de ecuaciones diferenciales.

Sistemas lineales de ecuaciones en diferencias y potencias de una matriz. Comportamiento a largo plazo de las soluciones: estabilidad, estabilidad asintótica e inestabilidad.

Matrices estocásticas y cadenas de Markov: estabilidad y vectores estacionarios

Análisis de series temporales.

Estudio de los sistemas dinámicos.

Tales tipos de aplicaciones y muchas otras más, hacen del problema de autovalores una técnica básica con aplicación en todas las ramas científicas, entre ellas la economía.

Palabras clave: autovalores, autovectores, QR, Schur, Jacobi, Hessenberg, Lanczos, Arnoldi, Cuppen, Krylov, Rustishauser

1 Introducción

El estado actual de los métodos para resolver el cálculo de los autovalores de una matriz es el resultado continuado de pequeños pasos. En el presente escrito se van a describir todas aquellas ideas que aun siguen en vigor contribuyendo en el desarrollo de potentes algoritmos numéricos, fruto de la investigación del problema durante el siglo XX.

El problema a tratar es el de las soluciones no triviales de la ecuación $Ax=\lambda x$, donde A representa una matriz cuadrada nxn, λ es un autovalor y x un autovector asociado a λ

Teóricamente, el problema es equivalente a resolver: $\det(A-\lambda I)=0$ dando lugar al polinomio característico cuyas raíces son los autovalores buscados. Tal estrategia sólo puede seguirse en casos muy especiales, pues dicho método es muy inestable; pequeños cambios en los coeficientes del polinomio dan lugar a grandes cambios en las λ_i.

Dicho planteamiento sirve, sin embargo, para darnos cuenta de que, al contrario que ocurría con la resolución de sistemas lineales, todo método utilizado para hallar los autovalores de una matriz general debe ser de naturaleza iterativa, ya que el teorema de Abel-Ruffini nos demuestra la imposibilidad de hallar las raíces de un polinomio general de grado mayor que 4 mediante un algoritmo en un número finito de pasos.

Esto no quiere decir que no podamos dar respuesta al problema, pues se han desarrollado numerosos algoritmos numéricos que calculan con gran aproximación los autovalores y autovectores cuya velocidad de convergencia es rápida.

La elección de un algoritmo u otro, se realiza considerando aspectos sobre la estructura y propiedades de la matriz A (si es simétrica, real, Hermitiana, hemisimétrica, unitaria, densa, rala,) y según los resultados que busquemos (autovalores de mayor módulo, o los de menor módulo, la parte real de los autovalores negativos, si deseamos también los autovectores...).

Una vez abordado el problema inicial, las soluciones numéricas halladas arrojan luz sobre problemas más generales: $Ax=\lambda Bx$, y el problema de autovalores cuadrático: $Ax+\lambda Bx+\lambda^2 Cx=0$.

2 Formas canónicas.

Una manera de abordar el problema es intentar reducirlo a otro problema equivalente en el cual su estructura es más simple. Así, mediante transformaciones, la matriz A puede reducirse a una forma canónica.

Toda matriz real y simétrica A es semejante a una matriz diagonal D y la matriz de paso es ortogonal $Q^T AQ=D$. A y D poseen el mismo conjunto de autovalores y autovectores

Las matrices no simétricas en ciertos casos pueden diagonalizarse encontrándose una matriz X no singular (en general no ortogonal) tal que $X^{-1}AX=D$ (si la multiplicidad de cada autovalor λ es igual a la dimensión del subespacio propio asociado a λ). En particular podremos, si los autovalores son distintos, pues su multiplicidad es 1.

En todo caso toda matriz se puede llevar a la forma de Schur $Q^*AQ=R$ con Q ortogonal y R triangular superior. Siendo el caso simétrico una particularización ya que una matriz simétrica triangular es diagonal.

Otra forma canónica es la forma de Jordan, aunque el cálculo de ésta resulta altamente inestable, ya que una pequeña perturbación conduce a una matriz con autovalores muy diferentes. El equivalente a la forma de Jordan para el problema generalizado $A-\lambda Bx=0$ es la forma de Kronecker.

3 Teoremas de Perturbación

Muchos algoritmos consisten en reducir la matriz A a una forma canónica como acabamos de ver, por ejemplo una matriz diagonal o triangular. Dicha reducción se realiza de forma iterativa y hay que presentar criterios para considerar que la matriz está suficientemente próxima a la forma final, en cuyo caso pararemos. Estos criterios se basan en los teoremas de perturbación.

Uno de los resultados más utilizados es el teorema de los discos de Gerchgorin, que nos dice que los autovalores de una matriz $A=(a_{ij})$ están localizados en la unión de círculos con centro a_{ii} y radio $\Sigma_{j\neq i} |a_{ij}|$; otro teorema afirma que si un número s de esos discos forman una componente conexa aislada de los demás, entonces este grupo contiene s autovalores y en particular si un sólo disco está aislado, en él se encuentra un único autovalor.

Otros teoremas se basan en que los autovalores varían de manera continua con los valores de los elementos de A, y si λ tiene multiplicidad uno, ésta dependencia es diferenciable.

Weyl probó que si los autovalores de δA son: $\eta_n \leq .... \leq \eta_2 \leq \eta_1$ entonces los autovalores γ_i de $A+\delta A$ cumplen: $\lambda_k + \eta_n \leq . \gamma_k \leq . \lambda_k + \eta_1$.

La propiedad de interlazado es muy utilizada en algoritmos de tipo proyectivos, siendo A_r el menor principal de A de orden rxr con autovalores λ^r_j.

$$\lambda_{r+1}^{r+1} \leq \lambda_r^r \leq \lambda_r^{r+1} \leq ... \leq .\lambda_2^{r+1} \leq \lambda^r_1 \leq \lambda_1^{r+1}$$

Otro resultado importante es utilizado en el método de “divide y vencerás”, que conlleva perturbaciones de rango 1 (se expresa la matriz A como suma de dos submatrices de orden menor y una corrección de rango 1). Si $B=A+\tau cc^T$ con $||c||^2 = 1$, $\tau \geq 0$ real, entonces los autovalores de B $\lambda_i(B) \in [\lambda_i , \lambda_{i-1}]$ y existen $\alpha_1,...,\alpha_n$ siendo $\Sigma_i \alpha_i=1$ tal que $\lambda_i(B)=\lambda_i+\alpha_i \tau$

Un resultado del teorema de Kato-Temple obtiene cotas para los cocientes de Raleigh para matrices simétricas en función del gap (salto) gap(i,A)=min $j\neq i$ $|\lambda_j-\lambda_i|$.

4 Método de Jacobi

¡Es un método que data de 1846!. Consiste en ir aproximando la matriz original real y simétrica a una matriz diagonal por una iteración de sucesivas rotaciones planas. Dichas rotaciones se van aplicando al elemento de mayor módulo fuera de la diagonal, de tal manera que lo anule. En cada paso pueden destruirse algunos ceros anteriormente conseguidos, pero el efecto total es la reducción de la magnitud de los elementos no nulos.

Debido al alto coste de encontrar en cada paso el elemento de mayor magnitud, el algoritmo se corre de manera cíclica tratando en orden los elementos a anular.

Cada rotación afecta únicamente a un par de filas y columnas de la matriz A por lo que el método es altamente paralelizable; su otra gran ventaja es la gran precisión de los autovalores aproximados hallados, aunque su contrapartida es la lentitud de su convergencia.

Se han hecho intentos de generalizar el método a matrices no simétricas hallando ceros para conseguir la forma de Schur, pero para este caso otros métodos resultan más adecuados. El método QR puede verse como de tipo Jacobi, pues produce la forma de Schur mediante una serie de transformaciones de semejanza compuestas de rotaciones.

5 Método de las potencias

En su forma primitiva, tiene la ventaja de su gran simplicidad. Consiste en tomar un vector adecuadamente elegido con componente no nula en el vector v_0 correspondiente al autovalor de módulo mayor; tras multiplicaciones sucesivas por la matriz A del vector resultante, estos vectores van adquiriendo mayor peso en la dirección del autovector $v_{0.}$ Como veremos en el siguiente apartado, estos vectores definen los subespacios de Krylov.

La limitación del método así expuesto, es que resulta sólo válido para el cálculo del autovalor de módulo mayor y su velocidad de convergencia es lineal con el cociente entre el primer autovalor y el segundo, resultando muy lenta si el cociente es próximo a uno.

Wielandt propuso el método de la deflación para eliminar el autovalor mayor una vez hallado y continuar el proceso; ésta es la deflación implícita. También se puede realizar de manera explícita aplicando el método de potencias a un segundo vector ortogonal al subespacio de autovectores ya hallados.

La iteración inversa consiste en aplicar el método de potencias a la matriz $(A-\mu I)^{-1}$ siendo los vectores propios de esta matriz los mismos que los de A.

Resolvemos por medio de un sistema de ecuaciones lineales $(A-\mu I)^{-1}w= v^{k-1}$ y luego normalizamos $v^k=w/||w||$ y seguimos iterando.

El método converge a pesar del mal condicionamiento de la matriz perturbada si μ se aproxima a λ, resultando el método muy potente para hallar los vectores propios de A.

Mediante el cociente de Rayleigh $\lambda^k=[v^k]^T A\ v^k$ (v^k normalizado), obtenemos una buena aproximación de un valor propio si v está próximo a un vector propio. La iteración del cociente de Rayleigh consiste en aportar una aproximación del valor propio por medio del cociente de Rayleigh, para seguidamente aplicar $(A-\mu I)^{-1}$ hallando una aproximación del autovector propio más próximo al autovalor anterior. Ostrowski estableció convergencia cúbica bajo ciertas circunstancias tanto para el caso simétrico como el no simétrico.

La Treppeniteration (iteración en escalera) consiste en multiplicar la matriz A no por un sólo vector sino por una matriz de vectores L_s (nxs), factorizando el resultado por transformaciones gaussianas como L_{s+1} R_{s+1}. Si los autovalores son distintos, R_{s+1} converge a una matriz cuyos autovalores son los del subespacio dominante.

Rutishauser observó que si factorizamos A=LR, entonces $L^{-1}AL= L^{-1}LRL=RL =A_1$

Volviendo a factorizar $A_1 = L_1R_1$ e iterando, da lugar al método LR. Si en lugar de una matriz triangular inferior utilizamos una matriz ortogonal Q resulta el potente método QR.

La Combinación del método de potencias para un número de vectores menor de n junto con la ortogonalización QR da lugar al método de iteración simultanea.

6 Algoritmos de reducción

La estrategia de este tipo de algoritmos consiste en realizar una primera fase que obtenga una matriz semejante a A, con una estructura más sencilla, por ejemplo, tridiagonal en el caso simétrico y en forma de Hessenberg (con ceros debajo de la subdiagonal), esto puede realizarse en un número finito de pasos. La segunda fase del algoritmo consiste en la iteración (por ejemplo pasar de la forma de Hessenberg a triangular) que en principio no termina, aunque en la práctica se consigue que converja en pocas iteraciones.

La ventaja del planteamiento se obtiene si al aplicar las iteraciones se conserva la forma reducida (por ejemplo vimos en el método de Jacobi que esto no sucedía, pero el método QR si conserva la forma de Hessenberg) lo cual garantiza la brevedad de los cálculos.

Para conseguir la reducción de la matriz pueden utilizarse rotaciones de Givens, o bien reflexiones de Householder.

Otros algoritmos de reducción se fundamentan en el subespacio de Krylov, generado por el método de las potencias visto anteriormente utilizando los vectores x, Ax, A^2x ...,

y haciendo uso del teorema de Caley-Hamilton que nos dice que toda matriz A satisface su polinomio característico, mediante los vectores del subespacio de Krylov podemos intentar reconstruir los coeficientes de dicho polinomio, aunque tal método corre el riesgo de ser muy inestable pues ya hemos apuntado que pequeñas variaciones en los coeficientes introducen grandes variaciones en las raíces del polinomio.

Además debido a que los vectores del método de potencias se aproximan los autovectores dominantes, el subespacio de Krylov resulta ser una base mal condicionada.

Hessenberg propuso utilizar una base de Krylov modificada, ortogonalizar a una base de pruebas, por ejemplo la base canónica. La misma idea es recogida por Lanczos para matrices simétricas y Arnoldi para las no simétricas, Ellos utilizan como base de pruebas la misma base de Krylov, ortogonalizando cada nuevo vector con el subespacio de Krylov hallado hasta el momento

También intentó Lanczos un método para matrices no simétricas construyendo dos bases de vectores una para el subespacio de Krylov de A y otro para el de A^T. Sin embargo el método resultó poco satisfactorio y fue Arnoldi quien introdujo un método de reducción para el caso asimétrico, ortogonalizando cada nueva base con todas las anteriores para conseguir una matriz semejante a A con la forma de Hessenberg.

El algoritmo QR es de enorme importancia debido a su rápida convergencia y estabilidad.
Fue descubierto independientemente en 1961 por Francis y Kublanovskaya, basado en el algoritmo LR de Rustishauser como anteriormente se mencionó.

Fue Francis quien encontró que en cada iteración el algoritmo, conserva la forma de Hessenberg, lo cual simplifica los cálculos a la vez que añade estabilidad.

La utilización de desplazamientos hace del algoritmo QR muy efectivo poseyendo orden de convergencia cúbico para matrices simétricas y cuadrático para las asimétricas.

Tras introducir dichos desplazamientos, la matriz deja de tener la forma de Hessenberg por lo que hay que utilizar una técnica de reducción para volver a ella; a estos métodos se le conoce como “bulge-chasing” (cazar la perturbación).

De principios de los años 60 data el método de la bisección estudiado por Givens, que consiste en hallar los ceros del polinomio característico, aunque sin hallarlos explícitamente sino mediante acotaciones decrecientes de los intervalos donde se hallan los ceros. Para el cálculo de dichos intervalos se utiliza la propiedad de que el número de autovalores negativos es igual al número de cambios de signo en la sucesión de Sturm (sucesión de determinantes de las submatrices principales de orden creciente). Haciendo uso de la propiedad de interlazamiento ya vista en el apartado 3 y de desplazamientos del origen, conseguimos acotar los autovalores en intervalos arbitrariamente pequeños.

Para que el cálculo de la sucesión de Sturm sea rentable, previamente se utiliza un método de reducción para obtener una matriz tridiagonal y los determinantes de las matrices principales están relacionados por una ecuación de tres términos.

En 1981 Cuppen propuso un algoritmo de tipo “divide y vencerás”, para matrices densas y simétricas. Su mecanismo se basa en que una matriz tridiagonal puede escribirse como la suma de una matriz diagonal por bloque y una corrección de rango uno.

El rango del problema se reduce en cada iteración a la mitad, más una corrección de rango uno cuyo cálculo no es lineal aunque si rápido llevando a una ecuación secular.

Es especialmente importante en este método las distintas correcciones debidas a Dongarra, Sorensen y más tarde Demmel, con el fin de solucionar problemas de estabilidad. Hasta hoy, constituye la opción más rápida para matrices simétricas de orden >25.

7 Métodos iterativos

Aunque teóricamente el método de Lanczos es finito, su puesta en práctica lleva a malos resultados debido a los errores de redondeo, aunque Paige mostró que podía ser viable si se utilizaba de manera iterativa.

Solucionando problemas de ortogonalización de los vectores, y estudiando el error en las aproximaciones a los autovalores (valores de Ritz) resultó un algoritmo potente para el caso de matrices simétricas ralas.

Del mismo modo se hicieron esfuerzos para mejorar el método de Arnoldi para el caso asimétrico. Y Davidson introdujo un método para matrices diagonalmente dominantes que luego seria la base para el algoritmo Jacobi-Davison.

8 Epílogo

Para concluir haremos un resumen del estado actual de los problemas de autovalores.

Para matrices reales y simétricas, el problema se resuelve fácilmente pues en este caso tenemos un sistema ortonormal de autovectores y los autovalores son reales, estas propiedades lo explotan los diversos métodos.

Para matrices simétricas de orden pequeño $n<25$, tenemos el algoritmo QR, si es mayor de 25 pero menor de unos cuantos miles, utilizamos la combinación del algoritmo "divide y vencerás" con el QR. Para matrices de orden mayor se utiliza el algoritmo de Lanczos.

Para el caso no simétrico, la solución no es tan clara, los métodos más estables se basan en pasar a la forma de Schur. Si el orden es pequeño suele utilizarse QR; para ordenes mayores la elección puede ser entre el método de Arnoldi o el de Jacobi-Davison pero aún queda mucho trabajo por investigar.

9 Referencias bibliográficas

V. Bargmann, C. Montgomery, J. von Neumann, Solution of linear systems of high order, Technical Report, Princeton, Institute for Advanced Study, 1946.

W.E. Arnoldi, The principle of minimized iteration in the solution of the matrix eigenproblem, Quart. Appl. Math.

Z. Bai, D. Day, Q. Ye, ABLE: An adaptative block Lanczos method for non-Hermitian eigenvalue problems, Technical Report Research REport 95-04, University of Kentacky, Lexington, KY, 1995.

Z. Bai, G.W. Stewart, SRRIT - A FORTRAN subroutine to calculate the dominant invariant subspace of a nonsymmetric matrix, Technical Report 2908, Department of Computer Science, University of Maryland, 1992.

Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, H. van der Vost, Templates for the Solution of Algebraic Eigenvalue Problems: a Practical Guide, SIAM, Philadelphia, PA, 1999.

Z. Bai, Progress in the numerical solution of the nonsymmetric eigenvalue problem, Number. Linear Algebra Appl. 2 (1995) 219-234.